IL23R Variation Determines Susceptibility But Not Disease Phenotype in Inflammatory Bowel Disease

CLINICAL–ALIMENTARY TRACT

IL23R Variation Determines Susceptibility But Not Disease Phenotype in Inflammatory Bowel Disease

MARK TREMELLING,* FRASER CUMMINGS,[‡,§] SHEILA A. FISHER,[‖] JOHN MANSFIELD,[¶] RHIAN GWILLIAM,[#] ANDREW KENIRY,[#] ELAINE R. NIMMO,[**] HAZEL DRUMMOND,[**] CLIVE M. ONNIE,[‡‡] NATALIE J. PRESCOTT,[‡‡] JEREMY SANDERSON,[‖] FRANCESCA BREDIN,* CARLO BERZUINI,* ALASTAIR FORBES,[‡‡] CATHRYN M. LEWIS,[‖] LON CARDON,[§] PANOS DELOUKAS,[#] DEREK JEWELL,[‡] CHRISTOPHER G. MATHEW,[‖] MILES PARKES,* and JACK SATSANGI[**]

**IBD Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, England; [‡]Gastroenterology Unit, Radcliffe Infirmary, and [§]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, England; [‖]Division of Genetics and Molecular Medicine, Guy's, King's College and St Thomas' School of Medicine, London, England; [¶]Faculty of Medical Sciences, University of Newcastle, Newcastle upon Tyne, England; [#]Wellcome Trust Sanger Institute, Hinxton, Cambridge, England; [**]Department of Gastroenterology, University of Edinburgh, and Gastrointestinal Unit, Western General Hospital, Edinburgh, Scotland; [‡‡]Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, United Kingdom; and [§§]Centre for Gastroenterology and Nutrition, University College, London, England*

See editorial on page 2045; CME quiz on page 1999.

Background & Aims: **Identification of inflammatory bowel disease (IBD) susceptibility genes is key to understanding pathogenic mechanisms. Recently, the North American IBD Genetics Consortium provided compelling evidence for an association between ileal Crohn's disease (CD) and the *IL23R* gene using genome-wide association scanning. External replication is a priority, both to confirm this finding in other populations and to validate this new technique. We tested for association between *IL23R* and IBD in a large independent UK panel to determine the size of the effect and explore subphenotype correlation and interaction with *CARD15*.** ***Methods:*** **Eight single nucleotide polymorphism markers in *IL23R* tested in the North American study were genotyped in 1902 cases of Crohn's disease (CD), 975 cases of ulcerative colitis (UC), and 1345 controls using MassARRAY. Data were analyzed using χ^2 statistics, and subgroup association was sought.** ***Results:*** **A highly significant association with CD was observed, with the strongest signal at coding variant Arg381Gln (allele frequency, 2.5% in CD vs 6.2% in controls [$P = 1.1 \times 10^{-12}$]; odds ratio, 0.38; 95% confidence interval, 0.29–0.50). A weaker effect was seen in UC (allele frequency, 4.6%; odds ratio, 0.73; 95% confidence interval, 0.55–0.96). Analysis accounting for Arg381Gln suggested that other loci within IL23R also influence IBD susceptibility. Within CD, there were no subphenotype associations or evidence of interaction with *CARD15*.** ***Conclusions:*** **This study shows an association between *IL23R* and all subphenotypes of CD with a smaller effect on UC. This extends the findings of the North American study, providing clear evidence that genome-wide association scanning can successfully identify true complex disease genes.**

It is widely recognized that knowledge regarding the genetic basis of inflammatory bowel disease (IBD) and other complex diseases will provide key insights into pathogenic mechanisms. It is this fact that has spurred efforts to identify disease susceptibility genes. Of the many complex diseases investigated using molecular genetic techniques, Crohn's disease (CD) is exceptional in that specific genetic variants unequivocally associated with disease susceptibility have been successfully identified.[1,2] Nonetheless, characterization of the unknown number of remaining CD genes is required to complete the picture and remains a priority.

CD is one of the 2 common and related forms of IBD, the other being ulcerative colitis (UC). Within the United Kingdom, they have a combined prevalence of approximately 4/1000.[3] Both are known to have a significant genetic contribution to their etiology, but this is stronger for CD than UC.[4] The epidemiologic evidence also suggests that CD and UC share some susceptibility genes. In 2001, fine mapping of a widely replicated linkage region on chromosome 16 led to the identification of *CARD15* as a major CD susceptibility gene, with mutations leading to dysregulation of innate immune pathways.[1,2]

Abbreviations used in this paper: CI, confidence interval; IL23R, interleukin 23 receptor; SNP, single nucleotide polymorphism.

0016-5085/07/$32.00
doi:10.1053/j.gastro.2007.02.051

CARD15 genes have subsequently been shown in meta-analysis to predominantly determine susceptibility to ileal CD. Variants within a number of other genes have been associated with CD, UC, or both,[5–9] although their exact roles in IBD susceptibility require clarification and, in some cases, replication.

To date, pinpointing of disease genes has depended on detailed evaluation of candidates implicated by their function or patterns of expression or by fine mapping within large regions identified in the course of genome-wide linkage scans. Across the range of common diseases, productivity of such approaches has been limited. Most complex disease genetic studies, including many in IBD, have been beset by poor reproducibility of results and slow progress in identifying disease genes. This has been attributed to a range of factors, some of the most important being the low resolution of sib-pair linkage analysis, use of inappropriate statistical thresholds for significance, and poor matching of controls due to population admixture.[10] One powerful new method for the identification of complex disease genes is genome-wide association scanning, genotyping large panels of affected individuals and appropriately matched population controls for hundreds of thousands of polymorphic markers across the genome and using appropriately stringent statistical thresholds for significance.[11] Within the past year, such studies have become technically and financially possible using sets of markers that capture most of the common variation across the genome using knowledge regarding human haplotype structure available from the International HapMap Project (http://www.hapmap.org).[12] Systematic whole-genome association studies, in comparison with the previous gold standard of linkage analysis, should provide substantially increased power and resolution for detection of complex disease susceptibility genes.[13]

Recently, the results of a 308,332-marker genome scan in a North American panel of 547 non-Jewish case patients with CD and 548 controls were reported. Case patients were selected as having ileal CD to reduce heterogeneity.[14] Three markers showed a highly significant association with CD, 2 of which were in *CARD15*. The third marker was a rare coding variant rs11209026c (1142G→A; Arg381Gln) found in the interleukin 23 receptor (*IL23R*) gene on chromosome 1 ($P = 5.05 \times 10^{-9}$). Nine other markers showed association with $P < .0001$ either within *IL23R* or in the intergenic area with the adjacent *IL12RB2* gene. Internal replication was achieved in the index study using both a Jewish CD case-control cohort (peak P value, 3.36×10^{-13}) and family-based methodologies, the latter in addition suggesting association with UC in a small non-Jewish cohort. This finding indicates that *IL23R* may have a general role in the etiology of IBD.[14]

The aims of the current study were to seek replication of the association between *IL23R* and IBD in a large independent North European cohort representing the full range of CD and UC phenotypes, examine in detail genotype-phenotype relationships, explore evidence for epistasis with the known CD susceptibility gene *CARD15*, and provide accurate estimates of disease risk for associated variants. Replication of the association in an independent cohort would serve 2 important purposes. First, it is key to confirming the veracity of the original finding and the applicability of these findings in populations outside North America. Further, strong independent replication of the key finding of one of the first published genome-wide association scans would provide proof of principle that this novel methodology can be used to identify risk variants for complex diseases.

Table 1. Demographic Details of 2877 Individuals With IBD Used in Case-Control Panel

	CD (n = 1902)	UC (n = 975)
Median age at diagnosis (*y*)	26	38.9
Gender (F/M)	1153/745	480/495
Smoking at diagnosis (%)		
Never	58.4	55.0
Ex	9.4	30.3
Current	32.2	14.7
Jewish ancestry (%)	1.75	1.9
Nonwhite (%)	2.25	3.25
Surgery (%)	61.8	
Location/extent (%)	32.7 ileal 31.8 colonic 35.5 ileocolonic 27.1 perianal	16.5 rectum only 35.0 distal to splenic flexure 48.5 proximal to splenic flexure
Behavior (%)	36.5 stenosing 17.15 penetrating	

Subjects and Methods

Subjects

A total of 2877 individuals with IBD (1902 with CD and 975 with UC) were recruited in 5 centers across England and Scotland. The study was approved by the research ethics committees at each center.

Standard clinical, radiologic, and histologic diagnostic criteria were applied.[15] Phenotypic details were obtained by retrospective case notes review. CD phenotype was classified by age at diagnosis, location, and behavior of disease. Only one member of multiply affected families was included. A total of 1.75% were of Jewish origin, and 2.25% were nonwhite. Demographic and subphenotype data are presented in Table 1.

Control allele frequencies were obtained from 1345 individuals recruited across Britain as part of the 1958 British birth cohort.[16] Cases and controls were categorized into 12 broad geographical regions within Great Britain to minimize confounding due to variation in allele frequencies across the country.[17]

Table 2. Case-Control Allele Frequencies and Disease Odds Ratios (95% Confidence Intervals) for CD and UC

SNP	Allele	Controls	CD	*P*	Odds ratio (95% CI)	UC	*P*	Odds ratio (95% CI)
rs1004819	T	0.307	0.383	1.1×10^{-8}	1.41 (1.23–1.56)	0.348	.00713	1.20 (1.05–1.37)
rs10489629	G	0.448	0.372	1.8×10^{-8}	0.73 (0.66–0.82)	0.43	.26	0.93 (0.82–1.05)
rs11465804	G	0.058	0.025	7.2×10^{-11}	0.41 (0.31–0.53)	0.046	.081	0.77 (0.58–1.02)
rs11209026	A	0.062	0.025	1.1×10^{-12}	0.38 (0.29–0.50)	0.046	.0291	0.73 (0.55–0.96)
rs1343151	T	0.332	0.266	1.1×10^{-7}	0.73 (0.65–0.82)	0.315	.26	0.92 (0.81–1.06)
rs10889677	A	0.315	0.398	3.4×10^{-10}	1.45 (1.28–1.61)	0.358	.0042	1.22 (1.07–1.39)
rs11209032	A	0.320	0.390	1.3×10^{-7}	1.35 (1.22–1.52)	0.3524	.032	1.16 (1.01–1.32)
rs1495965	G	0.447	0.517	3.4×10^{-7}	1.32 (1.19–1.47)	0.457	.57	1.04 (0.92–1.18)

Genotyping

Genotyping of cases was undertaken with iPLEX chemistry on a matrix-assisted laser desorption/ionization time-of-flight MassARRAY platform (Sequenom, San Diego, CA). Cases were genotyped for 8 *IL23R* markers reported in the index study, including the nonsynonymous single nucleotide polymorphism (SNP) rs11209026 encoding amino acid change Arg381Gln (primer sequences in Supplementary Table 1; see supplemental material online at www.gastrojournal.org). Two of the North American markers (rs7517847, rs2201841) were omitted due to their location within a sequence of interspersed low-complexity repeats.

Genotyping of controls was undertaken at the Wellcome Trust Sanger Institute using the Illumina 550K chip (Illumina, San Diego, CA). Concordance of genotype calls between the different platforms was confirmed by genotyping 87 control DNAs for all 8 markers using the MassARRAY platform with strong concordance of calls between technologies—98.99% for the 8 markers overall. There was 100% concordance for 3 markers, including the coding variant Arg381Gln (Supplementary Table 2; see supplemental material online at www.gastrojournal.org). The data for 1594 cases of CD genotyped for *CARD15* mutations in earlier studies were used to undertake analysis for evidence of interaction between *CARD15* and *IL23R*.[18–21]

Statistical Methods

Allele frequencies were compared between cases and controls and between phenotypic subgroups using χ^2 tests of 2×2 tables. Odds ratios were calculated for the minor allele at each SNP; confidence intervals (CIs) were calculated using Woolf's method.[22] Pairwise SNP linkage disequilibrium coefficients were estimated using Haploview.[23] Conditional association analysis was implemented using COCAPHASE, a module of the UNPHASED program.[24] This method tests for equality of odds ratios for haplotypes identical at conditioning loci. The Mantel–Haenszel test for association conditioning on geographical region was implemented using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). Median age at disease diagnosis between groups was compared using the Wilcoxon rank sum test. Age at diagnosis was dichotomized according to the Montreal classification.[25] Unless specified otherwise, all analyses were performed using R version 2.2 for Windows (http://www.R-project.org).

Results

All genotypes were in Hardy–Weinberg equilibrium in both cases and controls ($P > .05$). A highly significant association with CD was observed across the region (Table 2). The strongest association was observed at the nonsynonymous SNP Arg381Gln, where the frequency of the A allele was 2.5% in CD compared with 6.2% in controls ($P = 1.1 \times 10^{-12}$). The odds ratio for this protective allele was 0.38 (95% CI, 0.29–0.50). Alternatively, the common wild-type homozygous GG genotype can be considered as the risk genotype with an odds ratio of 2.70. To minimize potential confounding from regional differences in allele frequencies, a Mantel–Haenszel test was performed across 12 regional strata. Mantel–Haenszel odds ratios was very similar to those obtained from pooled data for all SNPs. For example, the Mantel–Haenszel odds ratio was 0.36 (95% CI, 0.25–0.51) for Arg381Gln.

Several SNPs also showed significant association with UC (Table 2). The strongest signal was observed with common SNPs rs1004819 ($P = .0071$) and rs10889677 ($P = .0042$). The frequency of Arg381Gln was only marginally different between cases and controls (UC, 0.046; controls, 0.062; $P = .029$), with an odds ratio of 0.73 (95% CI, 0.55–0.96). The nonsynonymous SNP Arg381Gln was in tight linkage disequilibrium with one other SNP (rs11465804, $r^2 = 0.85$) but weak linkage disequilibrium with all 6 other SNPs ($r^2 = 0.03$–0.1). A separate test for CD association was performed for each SNP conditioning on Arg381Gln by conditional regression modeling. This showed a significant association at all SNPs ($P < .001$) except rs11465804, with the strongest residual association detected at rs10889677 ($P = 4.6 \times 10^{-8}$). Hence, the nonsynonymous SNP does not account for all the association signal at this locus.

Data were then analyzed for evidence of significant genotype-phenotype correlations based on age at onset of CD, disease location, and disease behavior (Table 3). No

Table 3. Arg381Gln Genotype and Allele Frequencies in CD Cases Stratified by Known Phenotypic Subgroups and *CARD15* Status

	AA	AG	GG	Total	Freq(A)
Sex					
Male	1	36	690	727	0.026
Female	2	49	1068	1119	0.024
Smoking history					
No	1	28	729	758	0.020
Yes	1	21	414	436	0.026
Ex	0	1	127	128	0.004
Disease location					
Pure colorectal disease	1	23	515	539	0.023
Pure ileal disease	0	31	533	564	0.027
Ileocolonic disease	1	25	642	668	0.020
Any colorectal disease	2	46	1101	1149	0.022
Any ileal disease	1	54	1115	1170	0.024
Perianal disease					
Yes	2	20	454	476	0.025
No	1	59	1205	1265	0.024
Disease behavior					
Stenosing	1	32	604	637	0.027
Penetrating	2	15	248	265	0.036
Inflammatory only	0	30	714	744	0.020
Surgery					
Yes	2	51	1035	1088	0.025
No	1	31	645	677	0.024
Age at diagnosis (*y*)					
16 or younger	1	8	197	206	0.024
17–40	2	60	1129	1191	0.027
Older than 40	0	14	324	338	0.021
CARD15 status[a]					
−/−	3	52	1025	1080	0.027
−/+	0	15	342	357	0.021
+/+	0	5	98	103	0.024

[a]Samples are subdivided by *CARD15* status into those homozygous wild-type (−/−), those heterozygous for CD-associated variants (−/+), and those homozygous or compound heterozygous for CD-associated variants (+/+).

significant subgroup association was observed. In particular, the subgroup of subjects with CD affecting the colon only without small bowel disease (n = 539) appeared to be as strongly associated as those with exclusively ileal/small bowel involvement (n = 668) (minor allele frequencies, 2.3% and 2.0%, respectively). The age at disease onset ranged from 12 to 67 years in patients with CD who carried the A allele of Arg381Gln and from 0 to 80 years in wild-type GG cases. There was no difference in the median age of onset between these 2 groups (AA/AG: median, 28 years [n = 85]; GG: median, 26 years [n = 1650]; $P = .26$). Stratification of cases by age at diagnosis according to the Montreal classification[25] revealed similar genotype frequencies in all groups (Table 3). For UC, subgroup analysis by disease extent, smoking history, and sex also revealed no significant subgroup association. Age at onset of UC ranged from 14 to 79 years in cases who carried the A allele of Arg381Gln and from 2 to 81 years in wild-type GG cases, with no difference in the median age of onset between the 2 groups (AA/AG: median, 34 years [n = 72]; GG: median, 33 years [n = 708]; $P = .14$) (Table 4). A total of 1540 subjects with CD were fully genotyped for the 3 *CARD15* mutations (G908R, L1007fs, R702W) (Table 3). The frequency of Arg381Gln in 460 cases carrying at least one *CARD15* mutation (2.2%) was not significantly different from that in 1081 cases who carried none (2.7%; $P = .47$). None of the 3 cases who were homozygous for the rare A allele also carried a *CARD15* mutation.

Discussion

This study provides unequivocal confirmation of association between variants in the *IL23R* gene and IBD, suggesting a major effect on overall susceptibility to CD and a more modest effect on UC. Importantly, this study also shows the association at *IL23R* for the first time in a non-American population. The strength of this association at *IL23R* and the fact that it reaches such a magnitude in 2 independent data sets leaves no doubt that it is a true finding. In addition, this is one of the first instances of highly significant, independent replication of data derived from a genome-wide association scan and provides important validation of this technique as a hypothesis-free method for the identification of complex disease genes.

Table 4. Arg381Gln Genotype and Allele Frequencies in UC Cases Stratified by Known Phenotypic Subgroups

	AA	AG	GG	Total	Freq(A)
Sex					
Male	1	41	447	489	0.044
Female	2	43	436	481	0.049
Smoking history					
No	0	28	301	329	0.043
Yes	0	7	81	88	0.040
Ex	1	20	160	181	0.061
Disease extent					
Rectum only	1	12	134	147	0.048
Distal to splenic flexure	1	24	286	311	0.042
Proximal to splenic flexure	1	43	387	431	0.052
Age at diagnosis (*y*)					
16 or younger	0	2	39	41	0.025
17–40	1	41	443	485	0.044
Older than 40	0	28	228	256	0.055

As with the North American genome-wide scan, the strongest evidence for association was seen at the nonsynonymous SNP Arg381Gln, where the frequency of the A allele was 2.5% in CD compared with 6.2% in controls ($P = 1.1 \times 10^{-12}$). These allele frequencies are similar to those seen in the North American panel.[14] There was no evidence that *IL23R* variants associate with any particular subphenotype of CD based on disease behavior or location. Hence, there was no difference in minor allele frequency even between the extremes of pure ileal/small bowel CD and pure colonic CD (2.7% and 2.3%, respectively). Likewise, analysis based on disease behavior did not show any specific subgroup associations (Table 3). This negative result is interesting because it contrasts with the other confirmed CD susceptibility locus *CARD15*, which seems to have definite associations with ileal disease.[26] These findings are extended by the observation of association with UC overall but not with any known UC subphenotype group, suggesting that *IL23R* variants may exert a rather generic effect on chronic intestinal inflammation, although the effect size in UC does appear to be smaller than in CD. It is noteworthy that the odds ratio confidence interval at Arg381Gln for UC (0.73 [95% CI, 0.55–0.96]) does not overlap with that for CD (0.38 [95% CI, 0.29–0.50]), suggesting a significantly less marked protective effect of the rare allele for UC compared with CD.

Based on data from our large, independent panel of CD cases, it is possible to provide an accurate estimate of the size of the effect conferred by *IL23R* variants with regard to the risk of CD. We estimated an odds ratio of 0.38 (95% CI, 0.29–0.49) for Arg381Gln. This is likely to be a more accurate estimate than that provided in the index report from the North American study (odds ratio, 0.26; 95% CI, 0.15–0.43) due to the well-recognized bias of the so-called "winner's curse," which leads to overestimation of effect size in discovery panels.[27] Characterizing the exact effect size is important to permit sample size calculation for any further attempts at replication. Where the effect size is overestimated, there is a risk that apparently appropriately powered studies will fail to observe the effect and erroneously conclude that it is a false positive.

In some previous reports of genetic association, effect sizes have been quantitated as a population-attributable risk in addition to the odds ratio. This figure is intended to estimate the proportion of disease incidence attributable to a specific variant. However, it cannot be calculated for a protective minor allele as in the case of Arg381Gln. Further, while it is possible to think of the effect at Arg381Gln as an increased risk conferred by the common G allele, calculations of population-attributable risk based on this assumption lead to an implausibly high figure due to the very high carriage rate of the G allele in the control population.

One important question is whether the nonsynonymous variant accounts for all of the association signal at *IL23R*. This was tested by conditional regression modeling, looking for evidence of association while controlling for the effect at Arg381Gln. From this analysis, it is clear that there is a strong residual signal, maximal at rs10889677 ($P = 4.6 \times 10^{-8}$), and hence that variation at loci in addition to Arg381Gln, either within or adjacent to *IL23R*, exert an influence on IBD susceptibility. Whether this reflects a functional impact of the noncoding variants themselves or the fact that they are in linkage disequilibrium with other functionally significant or coding variants is yet to be established. Analysis within Haploview (http://www.broad.mit.edu/mpg/haploview/) of data available from the International HapMap Project[28] (http://www.hapmap.org) shows that this selection of 8 tag SNPs captures only 18 out of 83 informative SNPs, within *IL23R* and the 3′ intergenic region covered by these markers, genotyped in the CEPH (Utah residents with ancestry from northern and western Europe) panel at $r^2 > 0.8$. Any additional coding variation is likely to be rare because interrogation of Ensembl database release 41 (October 2006) (http://www.ensembl.org/) revealed the

presence of only 2 additional nonsynonymous coding variants (rs1884444 and rs7530511) in *IL23R* with minor allele frequency >1% in healthy European populations, both of which were investigated in the North American study with neither showing evidence of association. However, it is known that different splice isoforms of *IL23R* exist and it is possible that their expression is determined by some of the documented noncoding variation.[29] To clarify these issues, future studies will need to include resequencing of *IL23R* in a CD panel and fine mapping across the gene using markers identified as a result, as well as studies to assess the potential functional impact of variants identified.

The data with regard to Arg381Gln provide evidence of a very common variant being a disease risk allele, or conversely protection from CD being conferred by the rare allele. The explanations are likely to be complex but for immune-mediated conditions may include the fact that genetic variation at a particular locus confers a spectrum of risk, being protective against some diseases, such as infections, while increasing the risk of others, such as autoimmunity or inflammatory conditions. These variations will have been subject to differing selection pressures in diverse populations as a result of different environmental exposures. It is also noteworthy that for some of the markers showing evidence of association, it is the rarer allele that is associated with increased risk of CD. This further supports the argument for more than one variant in *IL23R* with different effects on gene function.

Recent studies have identified IL-23, the cognate ligand of *IL23R*, as a key player in both innate and adaptive immune systems. Most IL-23 is secreted by activated dendritic cells, monocytes, and macrophages following their exposure to pathogen-derived molecules that bind at toll-like receptors.[30] IL-23 stimulates a unique $CD4^+$ helper T-cell population characterized by the production of IL-17, tumor necrosis factor, and IL-6 and known as Th17 cells. These cells play a central role in driving autoimmune inflammation in a number of animal models. IL-17 stimulates monocytes and endothelial cells to produce proinflammatory mediators, which in turn promote rapid neutrophil recruitment.[30] The effect of IL-23 has recently been distinguished from that of the related heterodimer IL-12, with which it shares a common p40 subunit.[31] Importantly in this regard, 2 studies in knockout mice lacking the p19 subunit of IL-23 showed marked attenuation of T cell–mediated colitis, while knockout of the p35 subunit of IL-12 produces no such attenuation, suggesting that IL-23 but not IL-12 is essential for the development of colitis.[32,33] The identification of different roles for IL-12 and IL-23 in control of immune pathways together with the current genetic data suggest that targeting IL-23 (and components of its downstream effector pathway) may be a useful and specific strategy to inhibit IBD while sparing systemic host protective immunity.[34]

As well as focusing attention on the IL-23 pathway in the pathogenesis of IBD, the current study also provides key validation of genome-wide association scanning as a means of identifying complex disease susceptibility genes. The North American study group applied an appropriate, genome-wide significance level, and use of such a stringent threshold has immediately led to replication in our independent panel with a level of significance that makes the association indisputable.

To date, complex disease genetic studies have been beset by poor study design, particularly use of nonconservative thresholds for significance, resulting in publication of many unreplicated false-positive results across the spectrum of common disease, hence the importance of the current study in providing unequivocal early replication in an independent panel of the principal findings from one of the first reported genome-wide association scans. There are recent reports in another complex disease (age-related macular degeneration) that also provide grounds for optimism that this technique produces replicable genetic association data.[35–37] The clear message is that genome-wide association scanning works and that this study design, which is being so vigorously applied across a number of common diseases, is likely to be highly productive. The hope is that with use of appropriately stringent statistical thresholds and appropriately powered data sets, the success seen here in CD will be generally applicable without the plethora of false-positive results that have vexed the field of complex disease genetics to date.

Appendix

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1053/j.gastro.2007.02.051.

References

1. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603–606.
2. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599–603.
3. Stone MA, Mayberry JF, Baker R. Prevalence and management of inflammatory bowel disease: a cross-sectional study from central England. Eur J Gastroenterol Hepatol 2003;15:1275–1280.
4. Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of

monozygotic and dizygotic twins. A study of heritability and the influence of smoking. Gut 1988;29:990–996.
5. Reinhard C, Rioux JD. Role of the IBD5 susceptibility locus in the inflammatory bowel diseases. Inflamm Bowel Dis 2006;12:227–238.
6. Yamazaki K, McGovern D, Ragoussis J, Paolucci M, Butler H, Jewell D, Cardon L, Takazoe M, Tanaka T, Ichimori T, Saito S, Sekine A, Iida A, Takahashi A, Tsunoda T, Lathrop M, Nakamura Y. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. Hum Mol Genet 2005;14:3499–3506.
7. Ho G-T, Nimmo ER, Tenesa A, Fennell J, Drummond H, Mowat C, Arnott ID, Satsangi J. Allelic variations of the multidrug resistance gene determine susceptibility and disease behavior in ulcerative colitis. Gastroenterology 2005;128:288–296.
8. Franchimont D, Vermeire S, El Housni H, Pierik M, Van Steen K, Gustot T, Quertinmont E, Abramowicz M, Van Gossum A, Deviere J, Rutgeerts P. Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis. Gut 2004;53:987–992.
9. Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA, Rosenstiel P, Albrecht M, Croucher PJ, Seegert D, Nikolaus S, Hampe J, Lengauer T, Pierrou S, Foelsch UR, Mathew CG, Lagerstrom-Fermer M, Schreiber S. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet 2004;36:476–480.
10. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001;2:91–99.
11. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–1517.
12. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet 2006;38:659–662.
13. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95–108.
14. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee A, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH. A Genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006;314:1461–1463.
15. Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl 1989;170:2–6.
16. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol 2006;35:34–41.
17. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JMM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet 2005;37:1243–1246.
18. Waller S, Tremelling M, Bredin F, Godfrey L, Howson J, Parkes M. Evidence for association of OCTN genes and IBD5 with ulcerative colitis. Gut 2006;55:809–814.
19. Pearce AV, Fisher SA, Prescott NJ, Onnie CM, Pattni R, Green P, Forbes A, Mansfield J, Sanderson J, Schreiber S, Lewis CM, Mathew CG. Investigation of association of the DLG5 gene with phenotypes of inflammatory bowel disease in the British population. Int J Colorectal Dis 2007;22:419–424.
20. Arnott ID, Nimmo ER, Drummond HE, Fennell J, Smith BR, MacKinlay E, Morecroft J, Anderson N, Kelleher D, O'Sullivan M, McManus R, Satsangi J. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? Genes Immun 2004;5:417–425.
21. Ahmad T, Tamboli CP, Jewell D, Colombel JF. Clinical relevance of advances in genetics and pharmacogenetics of IBD. Gastroenterology 2004;126:1533–1549.
22. Woolf B. On estimating the relation between blood group and disease. Ann Hum Genet 1955;19:251–253.
23. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–265.
24. Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 2003;25:115–121.
25. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, Caprilli R, Colombel JF, Gasche C, Geboes K, Jewell DP, Karban A, Loftus Jr EV, Pena AS, Riddell RH, Sachar DB, Schreiber S, Steinhart AH, Targan SR, Vermeire S, Warren BF. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19(Suppl A):5–36.
26. Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JPA. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. Am J Gastroenterol 2004;99:2393–2404.
27. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003;33:177–182.
28. A haplotype map of the human genome. Nature 2005;437:1299–1320.
29. Zhang X, Zhang H, Zhang Y, Fu Y, He J, Zhu L, Wang S, Liu L. Identification and expression analysis of alternatively spliced isoforms of human interleukin-23 receptor gene in normal lymphoid cells and selected tumor cells. Immunogenetics 2006;57:934–943.
30. McKenzie BS, Kastelein RA, Cua DJ. Understanding the IL-23-IL-17 immune pathway. Trends Immunol 2006;27:17–23.
31. Oppmann B, Lesley R, Blom B, Timans JC, Xu Y, Hunte B, Vega F, Yu N, Wang J, Singh K, Zonin F, Vaisberg E, Churakova T, Liu M, Gorman D, Wagner J, Zurawski S, Liu Y, Abrams JS, Moore KW, Rennick D, de Waal-Malefyt R, Hannum C, Bazan JF, Kastelein RA. Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. Immunity 2000;13:715–725.
32. Yen D, Cheung J, Scheerens H, Poulet F, McClanahan T, McKenzie B, Kleinschek MA, Owyang A, Mattson J, Blumenschein W, Murphy E, Sathe M, Cua DJ, Kastelein RA, Rennick D. IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. J Clin Invest 2006;116:1310–1316.
33. Hue S, Ahern P, Buonocore S, Kullberg MC, Cua DJ, McKenzie BS, Powrie F, Maloy KJ. Interleukin-23 drives innate and T cell-mediated intestinal inflammation. J Exp Med 2006;203:2473–2483.
34. Schmidt C, Giese T, Ludwig B, Mueller-Molaian I, Marth T, Zeuzem S, Meuer SC, Stallmach A. Expression of interleukin-12-related cytokine transcripts in inflammatory bowel disease: elevated interleukin-23p19 and interleukin-27p28 in Crohn's disease but not in ulcerative colitis. Inflamm Bowel Dis 2005;11:16–23.
35. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385–389.
36. Edwards AO, Ritter R III, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. Science 2005;308:421–424.
37. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, Schnetz-Boutaud N, Agarwal A, Postel EA, Pericak-Vance MA. Complement factor H variant increases the risk of age-related macular degeneration. Science 2005;308:419–421.

Received December 14, 2006. Accepted February 8, 2007.

Address requests for reprints to: Jack Satsangi, FRCP, Gastrointestinal Unit, Division of Medical Sciences, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, England. e-mail: j.satsangi@ed.ac.uk; fax: (44) 0131-537-1085.

The authors acknowledge use of DNA from the 1958 British Birth Cohort collection (R. Jones, S. Ring, W. McArdle, and M. Pembrey), funded by Medical Research Council grant G0000934 and Wellcome Trust grant 068545/Z/02, the National Association for Colitis and Crohn's Disease and the Wellcome Trust for supporting the case DNA collections, and the Wellcome Trust Case Control Consortium, for which the Crohn's disease panel was originally assembled.

The authors have no conflicts of interest to declare.

M.T., F.C., and S.A.F. contributed equally to this work.

The authors thank all the subjects who contributed samples, as well as the consultants and nursing staff across the United Kingdom who helped with recruitment of study subjects: C. Todhunter, A. Sutherland, K. Mohiuddin, N. Thompson, M. Hudson, J. Barbour, P. Donaldson, S. J. Middleton, J. Woodward, J. Hunter, R. S. Harvey, J. H. Saunders, A. Douds, D. Sharpstone, S. Whalley, A. Nicolson, S. M. Greenfield, P. B. McIntyre, M. J. Carter, I. Barrison, H. J. Kennedy, I. W. Fellows, R. Tighe, M. G. Phillips, C. Jamieson, I. Beales, A. Hart, A. Prior, J. Wyke, S. Williams, Y. Miao, M. Ninkovic, M. Dronfield, P. Nair, R. Dickinson, P. Roberts, C. P. Willoughby, I. Dunkley, D. Morris, M. Twist, N. Fisher, D. Kelf A. Nightingale, C. W. Lees, G. T. Ho, I. D. Arnott, T. Ahmad, D. McGovern, J. Beckly, R. Cooney, L. Hancock, A. Geramia, S. Goldthorpe, and S. Patham.